Virtual Assistant Based on Recurrent Neural Networks: Scope and Coverage of Health Insurance

Diego Alberto Paz-Medina[1], David Eduardo Rojas-Cavassa[2], Ernesto Adolfo Carrera-Salas[3]
School of Engineering, Universidad Peruana de Ciencias Aplicadas (UPC), Lima, Peru [1, 2, 3]

***Abstract*—In Peru, the use of health services provided by state institutions has decreased, largely due to perceived deficiencies in care quality, such as delays in medical attention and administrative barriers that hinder timely access to information on insurance scope and coverage. This study develops a web-based application with an integrated chatbot that leverages Artificial Intelligence (AI) and a Recurrent Neural Network (RNN) to provide accurate and accessible information about institutional insurance, including the characteristics of different insurance types and specific coverage cases. The application integrates a chatbot powered by an AI model based on Natural Language Processing (NLP) from OpenAI, with intent recognition handled by a dedicated classifier. In comparison to existing chatbots in the healthcare field, the model proposes a hybrid approach based on RNN and GPT-3.5, optimized for health insurance queries in public sector contexts. The system was evaluated with 50 representative questions scored on four criteria: clarity, relevance, coherence, and accuracy. The chatbot achieved an overall mean response accuracy of 82% and a mean user satisfaction score of 4.59/5, indicating strong acceptance and usability. These results suggest that the combination of these technologies constitutes an effective alternative for addressing queries relevant to users of state health systems.**

***Keywords*—Insurance; insurance coverage; web application; chatbot; Recurrent Neural Networks**

I. Introduction

According to the nominal registry of insured individuals published in early 2022, reports that 98.96% of Peru's total population has health insurance coverage [1], with the majority covered by the Comprehensive Health Insurance (SIS) and Peru's Social Health Insurance (EsSalud), accounting for 69.5% and 29.3%, respectively. As stated in [2], in 2022, SIS showed a steady increase in enrollees from 2009 to 2022, while EsSalud exhibited a relatively constant affiliation trend. However, between 2020 and 2022, there was a decrease in the EsSalud-insured population, dropping from 30.4% to 22.8% of the total Peruvian population. In [3], the authors highlighted that, due to poor management and weak governance in recent years, the institution has experienced instability, resulting in delays in project implementation, decision-making processes, and turnover in key administrative areas. Additionally, [3] notes that the percentage of EsSalud-insured individuals visiting hospitals declined from 45% in 2004 to 31% in 2022. This decline is largely attributed to a perceived reduction in care quality [4], including delays in medical and consultation services, as well as the imposition of time constraints or administrative barriers that hinder effective understanding and access to consultation services.

Among the various approaches explored in previous research, [5] presents the comparison of a chatbot based on a Sequential Matching Framework (SMF), supported by a Recurrent Neural Network (RNN) algorithm, with other chatbots based on Natural Language Processing (NLP), such as RootyAI, ChatterBot, and DeepQA. In study [6], an artificial intelligence model was developed, capable of understanding terms related to the insurance domain for the resolution of queries. Although [7] and [8] document the development of a chatbot, [7] features the use of different types of algorithms, which are more advanced, such as the use of DeepQA for the chatbot's construction and DeepProbe for rewriting queries. In study [8], a chatbot was developed locally in hospitals, supported by Artificial Intelligence (AI), for the resolution of doubts related to immunization held by the medical personnel in charge of minors. In [9], the authors present an algorithm for adapting GPT models to specific tasks by combining information from different sources. This improves the accuracy of the model's responses and makes it more useful for specific tasks. This differs from the proposed approach, which focuses on the field of health insurance for the general public, where the user is not required to have prior knowledge of specific related terms. Finally, [10] describes the development of a chatbot in Ethiopia for mental health support, using a hybrid architecture based on Seq2Seq models, Amharic embeddings, and deep neural networks (LSTM and GRU).

In the present study, a web-based application was developed incorporating an integrated chatbot designed to provide both EsSalud-insured and uninsured individuals with access to information regarding insurance modalities, specific cases in which insurance coverage may apply, and healthcare centers within this public institution that offer treatment for the conditions specified by users. Based on the queries submitted by the user, the system generates contextually appropriate responses using an Artificial Intelligence (AI) model previously trained by a Recurrent Neural Network (RNN) algorithm. The model was trained in data related to the scope and coverage of the various insurance plans offered by EsSalud. Through the implementation of this application, users, regardless of insurance status, are provided with a digital channel through which frequently asked questions about EsSalud insurance can be addressed, thereby reducing waiting times for in-person assistance and alleviating congestion at insured service centers.

During the validation phase, the system's response accuracy was first assessed at the backend level. Specific validation criteria were established, and the system's responses were evaluated against official information obtained from EsSalud. The evaluation indicated an overall mean response accuracy of

82%, with the main limitation being the model's difficulty in generating clear and concise responses. Usability testing was then conducted with 32 participants who interacted with the chatbot, followed by a 10-item questionnaire designed to evaluate user experience. The results showed a general satisfaction score of 4.59 out of 5 on a Likert scale, a dropout rate of 16%, and a perceived overall improvement of 78% compared with human agents (41% rated the system as "much better" and 37% as "better"). In conclusion, the findings support the efficacy of digital solutions, such as AI-driven chatbots, in efficiently disseminating basic insurance-related information and responding to common inquiries, thereby improving service accessibility and responsiveness for both insured and uninsured populations.

The structure of the study is organized as follows: Section II provides a review of related works; Section III introduces the concepts applied throughout the development process; Section IV describes the methodology employed; Section V details the results of the experiment conducted; Section VI discusses the findings; and Section VII presents the conclusions.

II. Related Works

In [5], the authors present a study on the response-generation strategies employed by existing chatbots, aiming to identify shortcomings in maintaining coherent dialogues with users. Subsequently, a domain-specific chatbot named IntelliBot is proposed, which is based on a Sequential Matching Framework (SMF). This framework transforms each word into a vector according to its contextual meaning and employs hidden states from Recurrent Neural Networks (RNNs) to determine relationships between expressions. Additionally, the study examines alternative approaches for constructing conversational systems, including language generation techniques and neural network-based retrieval models. The model was trained and evaluated using three experiments and datasets. The first experiment utilized a movie dialogue dataset with over 200,000 samples and a batch size of 250 records. The second experiment used a credit card insurance dataset comprising over 10,000 samples with a batch size of 150 records. The third experiment involved a dataset of insurance-related terms and keywords with 1,000 samples and a batch size of 50 records. Following the execution of 71 questions across seven different categories, the results demonstrated the superiority of IntelliBot in terms of accuracy and memory, achieving a score of 0.97. In contrast, traditional chatbots obtained lower performance scores: 0.66 for RootyAI, 0.71 for ChatterBot, and 0.93 for DeepQA. However, despite these promising results, IntelliBot focuses primarily on general insurance contexts, without addressing the accessibility barriers or usability needs specific to public health institutions. Unlike the previous approach, the proposed model integrates a hybrid architecture optimized for state health insurance queries, user satisfaction, and accessibility for non-specialized users. This highlights a research gap regarding the adaptation of AI-driven conversational systems to public-sector health environments.

In study [6], an artificial intelligence model was developed using a Recurrent Neural Network (RNN) and a Convolutional Neural Network (CNN) to improve the resolution of queries related to the insurance domain. This model was created in response to the growing demand from individuals seeking information about the types of insurance they may obtain. Three datasets were used for training, validation, and testing, comprising a total of 16,889 questions, 16,889 correct answers, and 168,890 incorrect answers. The developed model demonstrated a higher accuracy rate than the baseline model, achieving a precision of 78.34%, which was attributed to its enhanced semantic understanding capabilities. This study addresses the same thematic area as the project under development, although it focuses on all types of insurance, whereas our work concentrates specifically on health insurance. Additionally, a key distinction lies in the response mechanism provided to users: while the referenced study focuses solely on the development of the model, the proposed project involves the implementation of the model within a digital alternative-specifically, a chatbot integrated into a web-based platform.

In [7], the authors present a design and implementation model for an intelligent chatbot using advanced algorithms. To achieve this, a high-quality dataset was required to train deep learning models effectively. The proposed model employed Long Short-Term Memory (LSTM) neural networks and the Sequence-to-Sequence (Seq2Seq) algorithm, along with the Bag of Words (BoW) model, beam search decoding for response prediction, and the BLEU-N algorithm to evaluate the chatbot's accuracy. As a result, the proposed method was compared and analyzed using the BLEU-N score against a traditional chatbot, a DeepQA-based chatbot, and rewritten versions of DeepProbe. The findings showed that the proposed method outperformed the others with a 59.44% improvement in BLEU-N score and overall performance. However, this approach focuses mainly on linguistic precision, without addressing user experience or applicability in public health contexts. The proposed model builds upon these advances by adapting deep learning methods to domain-specific insurance queries, helping to close this gap.

In study [8], an AI-assisted chatbot was developed to address questions related to childhood immunization in a low-resource and low-literacy setting in Karachi, Pakistan. The chatbot employed a continuous learning process through a Machine Learning algorithm integrated with a "Human-in-the-Loop" function, which aims to replicate natural human conversations without relying solely on keyword matching. Among the total number of inquiries (874), the most frequently addressed topics were vaccine expiration (32.4%), delays in vaccination due to child illnesses (16.8%), and management of side effects (15.7%). Additionally, it was found that most questions were answered promptly and in a manner that was clear to users. Regarding user acceptance, 90% indicated they would have no issue continuing to use the system; however, 40% expressed they would feel more comfortable knowing the source of the responses. Compared to the chatbot described in that study, the present research aims to address users' questions even when they have limited knowledge of the specific domain-namely, the scope and coverage of health insurance. This helps bridge the knowledge gap necessary for using a digital tool, as the user can make inquiries either in formal or colloquial language, both of which are interpreted in the same way thanks to the selected AI model.

In [9], the authors propose a method for adapting GPT-based artificial intelligence models to specific scenarios, such as patent claim generation, multilingual text summarization, and answer generation in question-answering systems. The new model, named Dempster-Shafer Theory (DST) framework, performs information fusion based on Dempster's theory of evidence, combining inputs from multiple sources to enhance accuracy and decision-making in response generation. The results showed that the proposed method improved the contextual accuracy of responses by 8.95% and 7.56%, respectively, compared to those generated by GPT-2 and GPT-3.5 models. However, its application remains focused on technical and text-processing domains, without exploring user-centered implementations in social or public service contexts. The present study addresses this limitation by applying GPT-based reasoning to healthcare insurance queries, emphasizing usability and accessibility.

In [10], the authors propose the design and implementation of an artificial intelligence-based chatbot called "PsychoBot", aimed at assisting users with potential mental health disorders in Ethiopia. For natural language processing, Word2Vec techniques (skip-gram and CBOW) were employed, along with Recurrent Neural Networks, including variants such as LSTM, GRU, double GRU, and transposed GRU. The architecture identified as optimal was the transposed GRU, due to its superior performance in terms of accuracy (79.86%), lower validation loss (0.208), and reduced training time (1434 minutes). In contrast to the present study, which focuses on the domain of health insurance, this work concentrates on psychological diagnostics, emphasizing the implementation in the local language and its robustness in handling diverse user queries. However, it does not integrate a web-accessible platform or a user-friendly chatbot interface, features that are essential to improving the usability of the developed solution and enhancing the overall user experience.

Table I presents a comparison of the methods used, datasets, and evaluation metrics across the related works.

TABLE I. COMPARISON OF RELATED WORKS AND THE PROPOSED MODEL

Year	References	Description of method/technique	Dataset size	Performance metric/s
2020	Paper 1 [5]	The working styles of chatbots are examined through a model based on Recurrent Neural Networks (RNN) and the Semantic Matching Framework (SMF).	71 questions across seven different categories.	Kappa coefficient and F1 score.
2023	Paper 2 [6]	The MFBT model, which focuses on insurance types, is developed using artificial intelligence with a BiLSTM-TextCNN architecture.	A total of 16,889 questions, divided into training, validation, and testing sets.	Model precision, recall, and F1 score.
2023	Paper 3 [7]	A design and implementation for an intelligent chatbot is presented, combining LSTM, a Seq2Seq architecture with attention mechanism, data preprocessing, Bag of Words (BoW) model, and supervised training.	Dialog dataset, containing 3,725 question–answer pairs.	BLEU score (Bilingual Evaluation Understudy).
2024	Paper 4 [8]	Bablibot: a chatbot designed to answer questions related to childhood immunization, employing a continuous learning process through Machine Learning with a "Human-in-the-Loop" function.	In Karachi (Pakistan), a thematic analysis of interviews was conducted with 2,202 caregivers, of whom 677 (30.7%) interacted with Bablibot.	User acceptance based on interview feedback.
2024	Paper 5 [9]	A method for adapting GPT models to various real-world scenarios to enhance performance in contextual dialogue classification.	Uses the Dialog dataset (3,725 Q&A pairs), DailyDialog (13,118 dialogues from everyday interactions), and WikiQA (3,047 questions sourced from Bing query logs).	BLEU score and cosine similarity score.
2025	Paper 6 [10]	A mental health support chatbot based on RNNs (LSTM, GRU), Seq2Seq architecture, Amharic word embeddings (Skip-gram and CBOW), and memory networks, combining generative and retrieval-based approaches in a hybrid architecture.	Employs a dataset compiled from the DSM-5 manual, client case histories, online psychological counseling websites, and general everyday conversations.	Precision, recall, and F1 score.
2025	The model proposed	A chatbot was developed integrating communication between an AI model and an RNN to address inquiries about the scope and coverage of EsSalud health insurance.	109 questions related to insurance services provided by EsSalud only for training purposes.	Average response accuracy rate and user survey results.

Source: Own Elaboration.

III. CONTRIBUTION

This section defines and contextualizes the key concepts underlying the proposed research. Each concept provides a basic definition and serves as a fundamental element for the design and implementation of the model.

A. Health Insurance

A contract established with an insurance company through the payment of a monthly premium. In exchange for the premium, the insurance company agrees to cover a variety of medical services, depending on the applicant's plan [11]. This concept was essential for structuring the chatbot's knowledge base. Based on the institutional documents collected, the system was trained to identify and respond to frequently asked questions regarding health insurance. A Recurrent Neural Network (RNN) was employed to accurately classify user queries, while the AI language model was responsible for generating clear and comprehensible responses. In contrast to previous chatbots that focused mainly on general healthcare information, the proposed model offers a domain-specific approach to health insurance, capable of delivering policy-based and context-aware responses aligned with institutional regulations.

B. Insurance Coverage

It refers to the protection provided by a health insurance policy against unexpected events, preventing these from becoming a significant financial burden in the present or future [12]. In the system's development, this concept was essential

for structuring the chatbot's responses regarding the services guaranteed by each type of insurance. Institutional documents in PDF format, which detail the coverage information, were used to fine-tune RNN algorithm for pattern recognition in user queries. Meanwhile, the OpenAI language model complements these responses by expanding the contextual information, helping users better understand which services are included in their health insurance plan. Unlike conventional systems with static information, the proposed model retrieves and integrates content directly from institutional documents to generate dynamic, personalized, and policy-specific responses, ensuring greater accuracy and relevance.

C. Insurance Scope

The limits of protection offered by an insurance policy in the event of unforeseen circumstances [13]. The concept of scope was important for training the chatbot to explain the limits and conditions of each insurance plan. This information, extracted from official documents, is processed by the RNN to accurately categorize user queries, while the AI model generates responses that clarify the extent of insurance coverage in various situations. Unlike traditional models based on predefined answers, the proposed system interprets each policy's scope dynamically, offering clearer, adaptive, and user-tailored explanations.

D. Artificial Intelligence

A scientific field focused on the development of computers and machines capable of thinking, learning, and making decisions in a manner similar to humans, or of processing data sets that exceed human analytical capacity [14]. The system is implemented using OpenAI's GPT-3.5 Turbo model, which interprets complex queries related to health insurance and generates coherent responses in natural language. This enables the chatbot to provide fast and accessible assistance without the need for human intervention. Unlike previous systems that used AI mainly for classification or retrieval tasks, this model employs generative AI to produce contextually accurate and user-friendly responses, improving accessibility for non-technical users.

E. Recurrent Neural Networks

A machine learning model based on neural networks that processes sequential data such as text or time series, converting it into specific sequence-based outputs [15]. They are used to process sequential health insurance queries, categorizing them based on user intent. A simple RNN architecture was employed due to its efficiency and speed in handling short sequence tasks, such as brief user queries. The aim is to provide a perspective on how this type of algorithm can "communicate" with others to enhance the responses it generates.

F. Web Application

A software program that runs directly on a web browser via the internet [16]. In the present work, this concept was applied when defining the appropriate medium for developing a solution that allows users to make inquiries about health insurance and access related information without requiring any external installation, by simply using their preferred web browser. Compared to the previously mentioned chatbots, the proposed solution aims to enable communication with the chatbot through an easily accessible web page, featuring a responsive design tailored to the user's needs.

IV. Method

This section presents the developed method according to the stages outlined in Fig. 1:

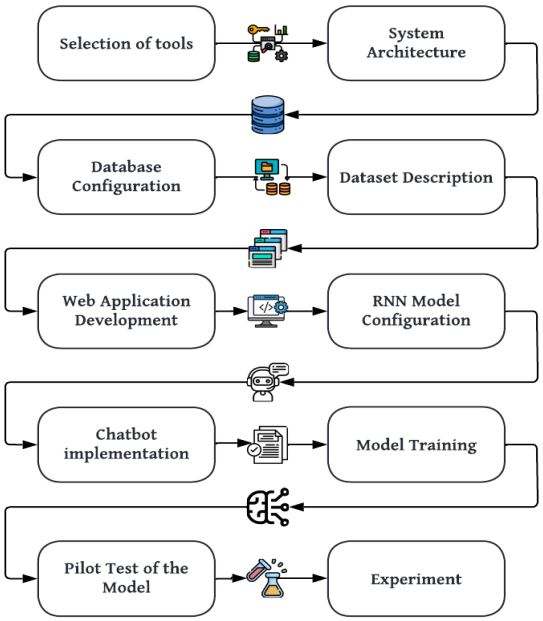

Fig. 1. Stages of the proposed method. Source: Own elaboration.

A. Selection of Tools

Regarding the AI provider, OpenAI was selected based on API maturity, tooling, and empirical evaluations of model quality reported in recent benchmarks [17]. GPT-3.5 Turbo was used for generating the chatbot's responses. Additionally, OpenAI was chosen for its widespread popularity in projects requiring human-like responses and for its ease of integration through the use of an API key, which grants access to OpenAI's models. Amazon Web Services (AWS) was chosen as the cloud service provider due to the broad range of services it offers for deploying web applications, machine learning resources, and its cost-effectiveness, making it the most suitable option considering the project's needs and constraints. Python was chosen as the programming language for data processing due to its robust ecosystem for NLP and machine learning, its ease of use, and its efficiency in handling large-scale datasets.

During the selection process for the most suitable AI model provider, four options were compared: OpenAI, Google Cloud AI, IBM, and Microsoft. The evaluation was based on five key criteria identified as most relevant in [18]: model governance, performance, ease of integration, scalability, and cost. OpenAI stood out for its high performance and scalability, being highly efficient in natural language processing tasks and capable of handling large data loads. Google Cloud AI provides robust tools but requires additional configurations to reach optimal performance. IBM facilitates model integration but tends to be more costly. Microsoft offers a wide range of options, though its implementation may require advanced knowledge and complex configuration.

Based on a qualitative analysis, the authors assigned scores to each provider according to the evaluation criteria. These scores were then weighted using a prioritization matrix that reflects the relative impact of each criterion. The results are presented in Table II, with OpenAI achieving the highest score of 3.94 as the most suitable AI model provider.

Following an evaluation of different database management systems, MySQL was selected due to its open-source nature. Unlike SQL Server, MySQL allows free usage and broad customization, as noted in [19]. It is recognized for its ease of installation and use, as well as its compatibility with multiple programming languages and operating systems. Moreover, it delivers optimal performance in data handling, enabling fast queries and updates. This makes it suitable for the current project, which does not require intensive performance, especially considering its lower resource requirements and ease of use without complex configurations. Additionally, its compatibility with Railway allows for seamless cloud deployment, accelerating testing and project management.

TABLE II. COMPARISON OF ARTIFICIAL INTELLIGENCE MODEL PROVIDERS

Criteria	*OpenAI*	*Google Cloud AI*	*IBM*	*Microsoft*
Model Governance	0.75	0.56	0.75	0.56
Performance	1.00	0.75	0.75	1.00
Ease of Integration	0.38	0.63	0.50	0.25
Scalability	1.25	0.75	1.00	0.75
Cost	0.56	0.56	0.56	0.56
Total	3.94	3.25	3.56	3.13

Source: Own Elaboration.

For this study, a simple Recurrent Neural Network (RNN) architecture was selected, thanks to its lightweight structure, quick training process, and suitability for short sequences-such as common questions related to health insurance. Unlike LSTM and GRU architectures, which are designed to capture long-term dependencies and process extended sequences, the simple RNN proved sufficient for our use case, where the inputs are brief and do not require long-term memory. While LSTM and GRU architecture are optimal for complex tasks, they involve a larger number of parameters, longer training times, and greater computational resource usage, as noted in [20]. Likewise, although Transformer models excel in complex tasks and large datasets, they involve high computational costs and require substantial data and specialized hardware, making them unfeasible for this study [21]. In contrast, the simple RNN offers a lighter and easier-to-implement design, which is essential to maintaining model simplicity without compromising its response generation capabilities. This comparison is shown in Table III, where an evaluation using a Likert scale from 1 to 5 was conducted. The Simple RNN achieved the highest overall score of 3.67, thus selected as the RNN architecture used in the proposed model.

TABLE III. COMPARISON OF RECURRENT NEURAL NETWORK ARCHITECTURES

Criteria	*Simple RNN*	*LSTM*	*GRU*	*Transformer*
Training Time	5	2	3	4
Long-term dependency capture	5	2	3	2
Number of parameters / complexities	1	5	4	5
Performance on short sequences	4	4	4	4
Training stability	2	5	4	4
Hardware requirements	5	2	3	2
Total	**3.67**	3.33	3.5	3.5

Source: Own Elaboration.

B. System Architecture

Fig. 2 presents the integrated system architecture, which follows a multilayer structure composed of the Presentation Layer, Application Layer, Data Layer, and Intelligent Service Layer; ensuring modularity, scalability and secure communication between all components.

Presentation Layer: Developed in Angular and hosted on AWS Amplify, this layer allows users to interact with the chatbot through a responsive web interface accessible from desktop and mobile devices. User queries are transmitted securely to the backend via RESTful APIs.

Application Layer: Implemented in Python (Django) and deployed on Railway Cloud, this layer processes user inputs, manages communication between the front end, the RNN algorithm, and the AI model. It also handles intent classification and message routing to ensure relevant and coherent responses.

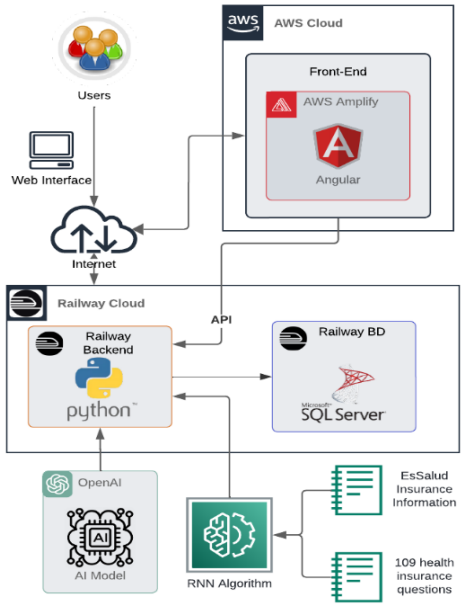

Fig. 2. System architecture of the proposed method.
Source: Own elaboration.

Data Layer: Hosted on Railway BD using Microsoft SQL Server, this layer stores insurance information, user interactions, and the dataset of 109 representative questions used for RNN training, enabling dynamic query retrieval and model improvement.

Intelligent Service Layer: Integrates the external OpenAI model with the internal RNN algorithm. The RNN identifies user intent, and OpenAI generates the final natural language response, combining deterministic classification with generative reasoning.

Interaction Flow: Users submit queries via the web interface; the backend classifies them using the RNN, retrieves related information from the database, and requests a response from the OpenAI model. The generated answer is then returned to the user interface, completing the interaction cycle efficiently.

C. Database Configuration

Since Railway was selected as the service provider, which allows integration with different database engines such as MySQL Server, this was used for storing the data related to the application. In addition, the same MySQL Server application was used to manage the data, both for queries and for editing. This, in turn, connects with the web application services to perform the queries required by the system.

A database model was developed using a class diagram and an entity-relationship model, having mainly the entities User, where personal data such as full name and login credentials will be stored, Responses, where the responses and the rating given by the user are stored to identify whether the response is accepted or rejected by the user in cases where more specialized attention is required, and Insurance, data related to the person's insurance, as well as the validity of the coverage.

D. Dataset Description

For the training and validation of the Recurrent Neural Network (RNN) model, a dataset was constructed from multiple sources containing information related to health insurance plans. Five PDF documents were used, each corresponding to a different type of health insurance. These documents included detailed information on the characteristics, coverages, exclusions, and validity periods of each insurance plan.

Additionally, a complementary PDF document containing 109 pairs of frequently asked questions (FAQs) and their corresponding answers was utilized. These questions were derived from real user inquiries and were validated and reviewed by personnel from the insurance department to ensure their accuracy and consistency. The questions were specifically designed to train the model in natural language understanding. To mitigate the limitations of the dataset, data augmentation techniques were applied, such as synonym replacement, paraphrasing, and back-translation, to increase linguistic variability while preserving semantic meaning.

Prior to model training, the question–answer pairs underwent preprocessing steps, including the removal of redundant symbols, text normalization (conversion to lowercase), and structuring into a suitable format for model input. The data were then tokenized, encoded, and used to train the RNN model. Through this process, the model learned to associate specific types of user queries with their appropriate responses based on the patterns within the dataset.

The outcome of this training provides a solid foundation for the chatbot to interpret user questions and generate coherent responses consistent with official information regarding the scope and coverage of health insurance plans.

E. Web Application Development

With respect to the visual layer for the user, it was decided to develop it using the Angular framework, given the development team's experience in programming business logic and modeling modules in TypeScript. This layer will be hosted on the AWS Amplify service due to its support for web applications, its ease of use for deployment, and its integration capabilities with other data processing services provided by AWS [22].

Its deployment will be carried out using Railway services, which will provide the functions performed in the backend through an API Gateway, which communicates with both the presentation layer on AWS Amplify and the artificial intelligence model provided by OpenAI [23]. To facilitate portability, scalability, and execution environment management, the backend has been developed and deployed within Docker containers, allowing for more efficient dependency control and greater consistency between development (localhost) and production (cloud) environments.

Communication begins with the front end in Amplify, sending data in JavaScript Object Notation (JSON) format to the backend, which is hosted on Railway services and receives the information through the API Gateway, then forwards this information to the OpenAI AI model [24].

F. RNN Model Configuration

The model architecture is based on a simple Recurrent Neural Network (RNN). Development and training were conducted in a local environment equipped with an Intel Core i5-6600 processor, 16 GB of RAM, and an AMD RX-480 dedicated graphics card. The system operated on Windows 10 Pro, using Visual Studio Code as the development environment, TensorFlow 2.15 as the deep learning framework, and Python 3.10 as the programming language.

During the experimental tests, equivalent SimpleRNN, LSTM and GRU models were trained under the same configuration parameters (number of neurons, embedding size, and sequence length). Although their accuracy was similar, the SimpleRNN achieved comparable performance with lower training time and memory usage, representing a more efficient cost–benefit balance for this system. Furthermore, since the input data do not involve long-term temporal dependencies, the use of LSTM or GRU would not significantly improve response quality, making the SimpleRNN a sufficient and lightweight solution.

The implemented architecture consists of three main layers: an Embedding layer, which transforms words into 64-dimensional numerical vectors; a SimpleRNN layer with 128 units, responsible for learning the sequential relationships

between words; and a dense output layer with a softmax activation function, which predicts the most probable response among the possible output classes.

Training was performed using the Adam optimizer with a learning rate of 0.001, employing sparse categorical crossentropy as the loss function and accuracy as the performance evaluation metric. The dataset was divided into 80% for training and 20% for testing, with a maximum sequence length of 20 tokens and a vocabulary size limited to 10,000 words.

The model was trained for 10,000 epochs, achieving stable convergence, and an adequate level of accuracy on the training data. Finally, the trained model, along with the tokenizer and label encoder, were stored in reusable formats, enabling their subsequent loading and integration into the web application.

The RNN model processes the user query through tokenization, padding, and intent classification, then integrates the result with the language generator. The full operational sequence is illustrated in Appendix A, which presents the pseudocode of the proposed model.

G. Chatbot Implementation

The application's front end, developed in AWS Amplify, was integrated with the backend services developed in Railway. Also, tests were conducted related to prompt construction to evaluate the request and response flow with the AI model. The backend services are hosted on Railway as the main server. For this purpose, Docker containers were used: one dedicated to the OpenAI-based chatbot, and another for the Recurrent Neural Network (RNN) model. Tests were conducted related to prompt construction to assess the request handling and the quality of the responses generated by the AI and RNN models.

Fig. 3. Chatbot view of the web application. Source: Own elaboration.

The main view of the developed chatbot is shown in Fig. 3. Users can submit queries regardless of whether they are registered, by entering a message in the input bar at the bottom, within the maximum character limit. Additionally, two options are provided for interacting with the chatbot. The first, "What type of insurance do I need?", displays a series of predefined questions and answers, which lead to a recommendation for one of the five available insurance plans offered by the institution, based on the user's selection. The second option, "Frequently Asked Questions", presents a list of stored questions and provides an answer upon selection, which is then sent to the endpoint for communication with the chatbot and the RNN algorithm.

H. Model Training

During the training phase, documentation related to the five types of insurance provided by EsSalud was collected, covering information about their coverage and the requirements for enrollment. These data were used to train both the RNN model and the external AI model provided by OpenAI. The main chatbot was built using language models from the same provider, responsible for generating dynamic and understandable responses to freely formulated user queries, with its responses strictly to the different types of health insurance. On the other hand, the RNNs were previously trained with specific data related to the insurance types. The objective of this model is to analyze and classify preconfigured or frequently asked questions (109 questions and answers), which require technical responses based on learned rules or patterns. The RNN serves as a specialized support system that enables fast and accurate responses without requiring as many resources as the main chatbot.

I. Pilot Test of the Model

Subsequently, a pilot test was conducted at the backend of the developed application, specifically focusing on the API used to communicate with both the RNN and AI models. For the test, a list of questions was prepared along with the corresponding answers obtained from EsSalud's insured services personnel. A total of 50 questions, along with their respective answers (distinct from the 109 used to train the model), were formulated and then compared with those generated by the developed model, using four criteria derived from those in [25]: Clarity (whether the response focuses on what is being asked without unnecessary detours), Relevance (whether the response is relevant to the context of the question), Coherence (whether the information is consistent with the question and logically coherent), and Accuracy (whether the response matches the verified information). Each criterion was assigned a weight of 25%, aiming to achieve an overall compliance rate above 80%. The comparison process was conducted by the development team, with support from personnel in the insurance systems department of the selected institution.

Table IV presents the results of the 50 questions posed to the AI model provided by OpenAI alongside the RNN model, evaluating the responses according to the established criteria.

In the responses obtained for the insurance-related questions, it was found that the answers were often not very clear, as they tended to include more information than necessary. For this reason, these questions had a compliance percentage ranging from 50% to 75%.

TABLE IV. CALCULATION OF THE RESPONSE ACCURACY PERCENTAGE OF THE AI MODEL

Questions	*Criteria*				*%*	*Questions*	*Criteria*				*%*
	Clarity	Relevance	Coherence	Accuracy			Clarity	Relevance	Coherence	Accuracy	
1	Yes	No	Yes	No	50%	**26**	Yes	No	Yes	No	50%
2	Yes	Yes	Yes	Yes	100%	**27**	Yes	Yes	Yes	Yes	100%
3	Yes	Yes	Yes	Yes	100%	**28**	Yes	Yes	Yes	Yes	100%
4	No	Yes	Yes	No	50%	**29**	No	Yes	Yes	Yes	75%
5	Yes	Yes	Yes	Yes	100%	**30**	Yes	Yes	Yes	Yes	100%
6	Yes	Yes	Yes	No	75%	**31**	No	Yes	Yes	No	50%
7	Yes	Yes	Yes	Yes	100%	**32**	No	Yes	Yes	No	50%
8	Yes	Yes	Yes	Yes	100%	**33**	Yes	Yes	Yes	Yes	100%
9	Yes	Yes	Yes	Yes	100%	**34**	No	Yes	Yes	Yes	75%
10	Yes	No	Yes	Yes	75%	**35**	No	Yes	Yes	No	50%
11	Yes	Yes	Yes	Yes	100%	**36**	Yes	Yes	Yes	Yes	100%
12	Yes	No	Yes	Yes	75%	**37**	Yes	Yes	No	No	50%
13	Yes	No	Yes	No	50%	**38**	Yes	Yes	Yes	No	75%
14	Yes	No	Yes	No	50%	**39**	Yes	Yes	Yes	No	75%
15	Yes	Yes	Yes	Yes	100%	**40**	Yes	Yes	Yes	Yes	100%
16	Yes	Yes	Yes	Yes	100%	**41**	Yes	No	Yes	No	50%
17	No	Yes	Yes	Yes	75%	**42**	Yes	Yes	Yes	No	75%
18	No	Yes	Yes	Yes	75%	**43**	Yes	Yes	Yes	No	75%
19	Yes	No	Yes	Yes	75%	**44**	Yes	Yes	Yes	Yes	100%
20	Yes	Yes	Yes	Yes	100%	**45**	Yes	Yes	Yes	No	75%
21	Yes	Yes	Yes	No	75%	**46**	Yes	Yes	Yes	Yes	100%
22	Yes	Yes	Yes	Yes	100%	**47**	Yes	Yes	Yes	Yes	100%
23	Yes	Yes	Yes	Yes	100%	**48**	Yes	Yes	Yes	No	75%
24	Yes	Yes	Yes	Yes	100%	**49**	Yes	Yes	Yes	Yes	100%
25	Yes	Yes	No	Yes	75%	**50**	Yes	Yes	Yes	Yes	100%

Source: Own Elaboration.

TABLE V. RESULTS OF THE AI MODEL EVALUATION

Value	
Mean	82%
Median	75%
Mode	100%
Standard Deviation	0.196
Minimum	50%
Maximum	100%

Source: Own Elaboration.

Table V presents the statistical data related to the responses provided by the artificial intelligence model alongside the RNN model. The mean accuracy was 82%, indicating a high level of acceptance, as it falls above the 70% threshold. The data show moderate dispersion around the mean, meaning the values are not highly spread out, but there is still noticeable variability. A mode of 100% was observed, making it the most frequent accuracy value among the responses. The answers fall within the upper range, with a minimum of 50% and a maximum of 100%, indicating that all responses had an accurate percentage above 50%.

A 95% confidence interval was calculated for the model's accuracy using the standard normal approximation, resulting in 0.82 ± 0.0543 (0.7657–0.8743). This interval confirms, with 95% confidence, the reliability and stability of the model's performance across the evaluated dataset.

J. Experiment

A test of the application was conducted with a group of 32 individuals who were either insured by EsSalud or interested in obtaining insurance from the institution. The users interacted

with the chatbot, powered by an RNN algorithm and an AI model, to make various insurance-related inquiries.

The participants were selected through a non-probabilistic sampling approach. The group consisted of 12 women and 20 men, all adults and active users of the national health insurance system. Participants were chosen based on their availability and familiarity with digital platforms, allowing an adequate representation of real insured users. The sample composition reflects the typical demographic distribution of the target population. This approach helps minimize potential bias and provides valid insights into user interaction with the proposed system.

Participants were then asked to complete a ten-question survey regarding their user experience, assessing their satisfaction with certain features or any difficulties encountered during usage. Table VI presents the scale and metric associated with each question.

TABLE VI. QUESTIONS ASKED IN THE USER SURVEY

No.	*Question*	*Scale*	*Metric*
Q1	How accurate do you consider the responses from the chatbot?	Likert	Average perceived accuracy by users
Q2	How often were the chatbot's responses useful and relevant to your inquiries?	Likert	Frequency of relevant vs. irrelevant responses
Q3	Are you satisfied with the time it takes for the chatbot to respond?	Likert	Satisfaction with response time and average response times
Q4	How easy was it to interact with the chatbot?	Likert	Average satisfaction with ease of use
Q5	What is your overall satisfaction level with the chatbot?	Likert	Average overall satisfaction
Q6	Did the chatbot effectively resolve your problem or answer your inquiry?	Dichotomic	Percentage of inquiries effectively resolved
Q7	Have you ever left the conversation with the chatbot before receiving a satisfactory response?	Dichotomic	Percentage of users who abandoned the conversation
Q8	Would you use this chatbot again in the future?	Dichotomic	Percentage of users with the intention to reuse
Q9	How would you compare the effectiveness of the chatbot to that of a human agent?	Likert	Perception of the chatbot compared to human support
Q10	How many times did the chatbot fail to understand or process your request?	Likert/ Intervals	Average number of errors or failures per session

Source: Own Elaboration.

V. RESULTS

Table VII shows the results obtained for questions Q1, Q2, Q3, Q4, and Q5, using a scale from 1 "Very Dissatisfied" to 5 "Very Satisfied". To calculate the average score per question, the sum of the products of each response value (1 to 5) and the number of users who selected that option was divided by a total of 32 users.

TABLE VII. RESULTS OBTAINED FOR QUESTIONS USING THE LIKERT SCALE

Result	*Q1*	*Q2*	*Q3*	*Q4*	*Q5*
5 - Very satisfied	21	20	23	25	20
4 - Satisfied	11	9	9	7	11
3 - Neutral	0	3	0	0	1
2 - Dissatisfied	0	0	0	0	0
1 – Very dissatisfied	0	0	0	0	0
Average	4.66	4.53	4.72	4.78	4.59

Source: Own Elaboration.

Regarding Accuracy of the Responses (Q1), 21 participants, representing 66%, indicated that they were Very Satisfied, and 11 participants, or 34%, responded that they were Satisfied. This indicates that the chatbot provides accurate and precise responses, with an average accuracy score of 4.66. The comparison is shown in Fig. 4.

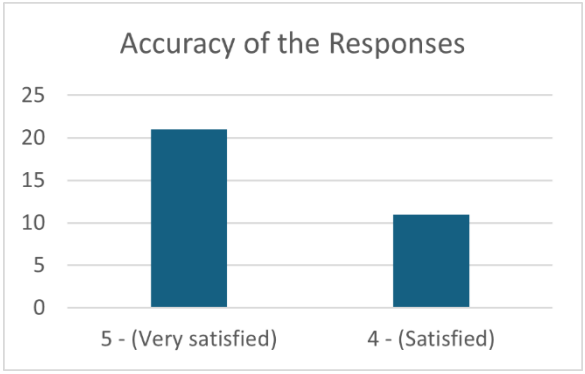

Fig. 4. Accuracy of the response results. Source: Own elaboration.

Regarding the Relevance of the Responses (Q2), 20 participants, representing 63%, indicated they were Very Satisfied; 9 participants, or 28%, responded that they were Satisfied; and 3 participants, representing 9%, expressed a Neutral opinion. This indicates that the chatbot is effectively providing relevant responses to users, with the majority expressing high satisfaction. The average frequency score was 4.53, as shown in Fig. 5.

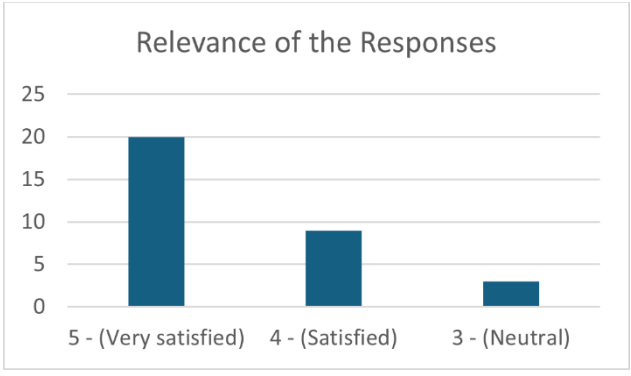

Fig. 5. Relevance of the response results. Source: Own elaboration.

Regarding Response Time (Q3), 23 participants (72%) indicated that they were Very Satisfied, and 9 participants

(28%) responded that they were Satisfied, leading to the conclusion that satisfaction with the chatbot's response time is adequate. The comparison is presented in Fig. 6.

Fig. 6. Response time results. Source: Own elaboration.

Regarding Ease of Use (Q4), it was found that 25 participants (78%) indicated they were Very Satisfied, and 7 (22%) responded that they were Satisfied. This indicates that the chatbot interface successfully provides a user-friendly tool, with an average satisfaction score of 4.78. The comparison is presented in Fig. 7.

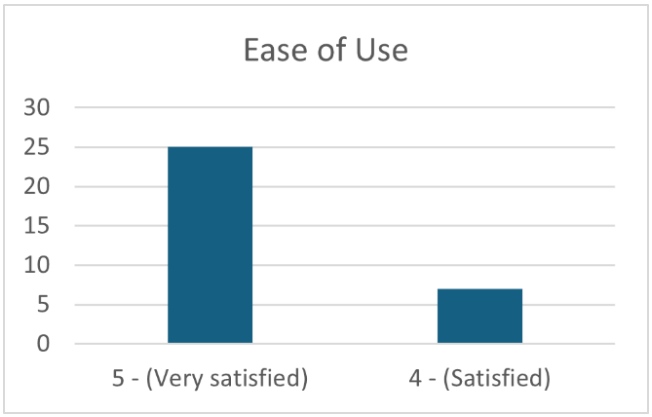

Fig. 7. Ease of use results. Source: Own elaboration.

Regarding the Overall Satisfaction Level (Q5), it was found that 20 participants (63%) indicated they were Very Satisfied, 11 (34%) responded that they were Satisfied, and 1 (3%) expressed a Neutral opinion. This indicates that the chatbot demonstrates a high level of overall satisfaction, with an average response score of 4.59. The comparison is shown in Fig. 8.

Table VIII shows the frequency of responses categorized into two options, 'Yes' and 'No', for the dichotomous questions.

Regarding the Question Resolution Rate (Q6), all participants (100%) perceived that the chatbot effectively resolved their submitted question, as shown in Fig. 9.

Regarding the Abandonment Rate (Q7), 5 participants (16%) indicated that they had abandoned the conversation before receiving a satisfactory response, while 27 participants (84%) did not abandon the conversation, resulting in a total abandonment rate of 16%. The comparison is shown in Fig. 10.

Fig. 8. Overall satisfaction level results. Source: Own elaboration.

TABLE VIII. RESULTS OBTAINED FROM DICHOTOMOUS SCALE QUESTIONS

ID	*Question*	*Yes*	*No*
Q6	Question Resolution Rate	32	0
Q7	Abandonment Rate	5	27
Q8	Intention to Reuse	32	0

Source: Own Elaboration.

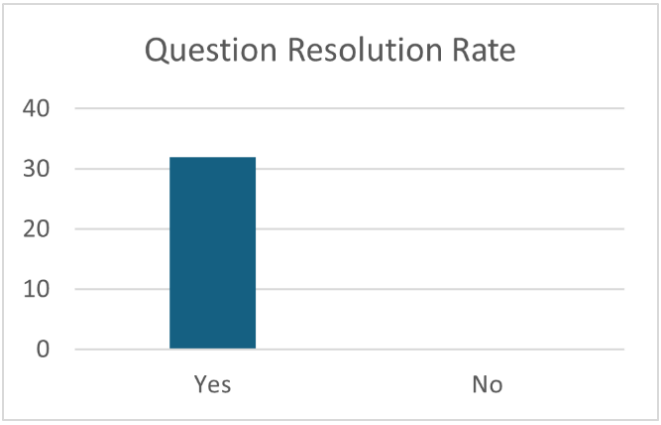

Fig. 9. Question resolution rate results. Source: Own elaboration.

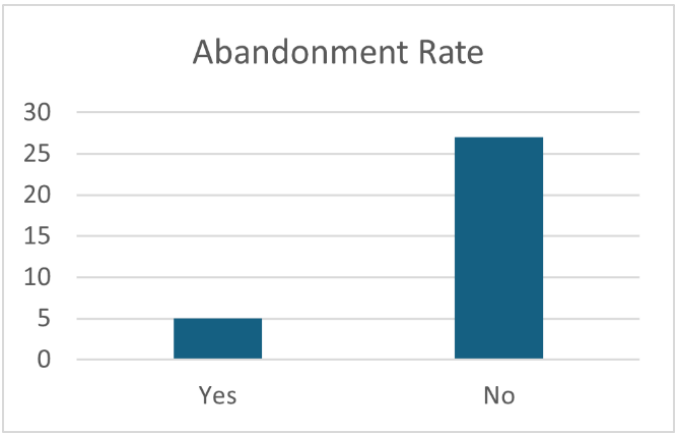

Fig. 10. Abandonment rate results. Source: Own elaboration.

Regarding the Intention to Reuse the Application (Q8), it was found that all participants (100%) would be willing to use this tool in the future. The comparison is shown in Fig. 11.

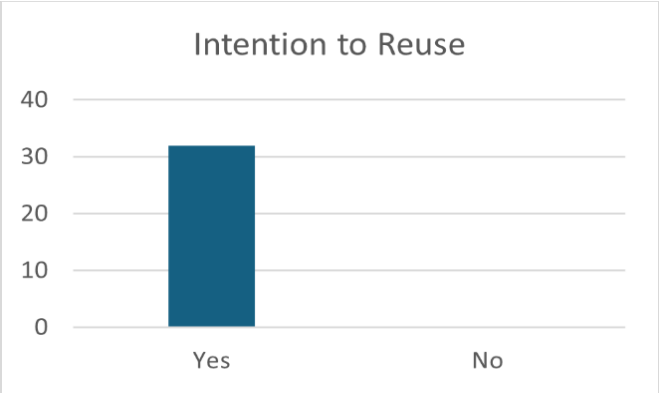

Fig. 11. Results on the intention to reuse. Source: Own elaboration.

Table IX shows the results obtained for the ninth question related to the comparison with human support, using a rating scale from "Much Worse" to "Much Better".

TABLE IX. RESULTS OBTAINED FOR QUESTION 9

Results	*Q9*
Much better	13
Better	12
Equal	7
Worse	0
Much worse	0

Source: Own Elaboration.

Regarding the Comparison with Human Support, 13 participants (41%) indicated that they consider the chatbot to be Much Better, 12 (37%) consider it Better, and 7 (22%) find it Equal to the attention provided by a human agent. The comparison is shown in Fig. 12.

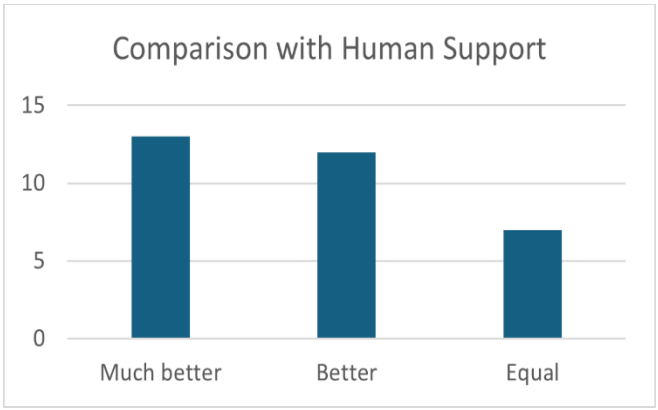

Fig. 12. Results on the comparison with human support. Source: Own elaboration.

Finally, Table X displays the results for question 10, focused on the number of errors encountered during queries.

Regarding the Error Rate (Q10), 20 participants (63%) reported not encountering any errors, 4 (13%) encountered 1 error, and 8 (25%) encountered 2 errors. There is an average of 0.63 errors found per user. The comparison is shown in Fig. 13.

TABLE X. RESULTS OBTAINED FOR Q10 REGARDING ERRORS

Results	*Q10*
0	20
1	4
2	8
3	0
4 or more	0

Source: Own Elaboration.

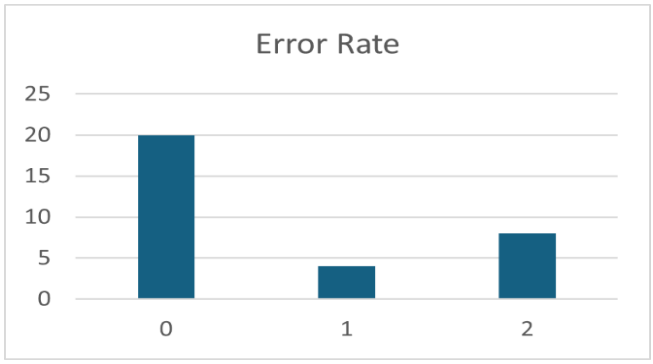

Fig. 13. Comprehension or processing errors encountered by users in the chatbot. Source: Own elaboration.

VI. DISCUSSION

The results of this study demonstrate that a chatbot integrated with artificial intelligence can be an effective solution to improve access to information about health insurance in Peru. Validation was carried out using 50 questions evaluated on the criteria of clarity, relevance, coherence, and accuracy, yielding an overall mean response accuracy of 82%, which, together with the usability findings, indicates that the chatbot achieved a high level of acceptance and reliable functionality.

In comparison with related work, study [5] employs an RNN along with a Sequential Matching Framework (SMF) that transforms words into vectors and analyzes complex relationships between expressions, achieving high precision (0.97). In contrast, the present work uses a simple RNN without SMF, focused on classifying frequently asked questions about health insurance. Instead of adding complexity, it is complemented by the OpenAI model, which effectively interprets context. This combination provides fast and clear answers to specific queries, reaching 82% compliance and 97% user satisfaction in testing.

In study [6], bidirectional LSTM neural networks (BiLSTM) are combined with a Text Convolutional Neural Network (TextCNN). Although both models rely on neural networks, the present study focuses on simple RNNs, which are more efficient for short questions and well-defined contexts such as EsSalud health insurance. This integration enables the system to respond to queries without the need for computationally intensive architecture or elaborate training processes, maintaining a good balance between performance and simplicity. In contrast, the hybrid BiLSTM-TextCNN approach focuses on recognizing general textual patterns, which are more complex.

On the other hand, study [7] employs advanced architecture that combines LSTM with a Seq2Seq model to develop highly complex chatbots, especially useful for extended dialogues. In contrast, this study opts for a simple RNN, avoiding the complexity of the Seq2Seq approach and focusing on brief, specific queries related to health insurance. This simplicity enhances efficiency while achieving similar results. Furthermore, the use of OpenAI's language model improves the understanding of question semantics and enables the generation of clearer and more appropriate responses. This makes the system more useful and practical when compared to models like DeepQA, which are more oriented toward maintaining long and complex conversations.

In study [8], the Bablibot chatbot was developed to answer questions related to immunization in Pakistan, utilizing a continuous learning process. The result was a 90% user acceptance rate. In comparison, the proposed model focuses on the domain of health insurance, an area in which users often have limited prior knowledge. Despite this, the proposed system achieved satisfaction rates of 63% (very satisfied) and 34% (satisfied), resulting in a total satisfaction score of 97%.

Additionally, study [9] employs Dempster-Shafer theory to fuse data from multiple sources, enhancing accuracy in complex and highly contextual questions. In contrast, the present work uses a simple RNN architecture without external data fusion, focusing on delivering immediate and accurate responses within the specific domain of health insurance. This narrower contextual scope enables faster response times and easier implementation.

Finally, in study [10], a variation of RNN known as Transposed GRU was used, which reverses the input data sequence during processing. In comparison, this study opted for a simple RNN to generate chatbot responses, which was likewise validated through user feedback via a survey. In the user validation, the present study achieved a perceived relevance of responses of 91% (63% "Very Satisfied" and 28% "Satisfied"), compared to the 76% obtained in the referenced study. Furthermore, the overall satisfaction level perceived by users was predominantly positive at 97% (63% "Very Satisfied" and 34% "Satisfied"), contrasting with the 76.2% acceptance rate reported in the referenced work.

Although the model demonstrated strong results, it has limitations in handling complex questions or extended dialogues due to the use of a simple RNN. In addition, depending on the OpenAI-provided AI model, constraints may arise in certain contexts. The validation was also conducted with a limited sample size, and expanding the study could provide more comprehensive feedback to identify critical improvement areas.

Despite these limitations, the model stands out for its simplicity, speed, and ease of implementation. These features make it potentially adaptable to other industries, such as education, public services, or other insurance providers, where quick and clear access to information is equally essential. Adjusting the content and context would be sufficient to address emerging community needs, thanks to the scalability of the system. This scalability depends on the databases to be incorporated for adaptation to other health domains, as well as the expansion of training questions to be used, thereby enriching the information possessed by the developed model.

VII. Conclusion

Artificial intelligence presents great potential to support customer service in the healthcare sector. Through an AI- and RNN-based chatbot, it is possible to automate repetitive tasks such as answering frequently asked questions and providing basic information about health insurance, thanks to the RNN algorithm's ability to select the most optimal response and the AI model's natural language generation capabilities. This technology not only improves response speed and reduces wait times but also demonstrates that intelligent automation can be integrated into both public and private healthcare institutions to enhance information accessibility and efficiency. Despite their relatively simple implementation, this type of solution demonstrates the feasibility of modernizing service channels using accessible and efficient tools, promoting a better experience for both insured and uninsured individuals, with the potential to enrich the chatbot's responses by increasing the training data for the RNN model or employing more specialized architectures tailored to this type of algorithm.

The evaluation results showed an overall mean response accuracy of 82% and high satisfaction levels across all criteria, indicating that the developed chatbot provides reliable, coherent, and contextually relevant responses. These findings suggest that combining an RNN classifier with a generative model like GPT-3.5 constitutes an effective hybrid strategy for managing health-related queries, achieving performance close to that of human support.

It was found that the developed chatbot and the designed interface were well received by users, with a majority of responses indicating "Very Satisfied" for Question P1 on perceived response accuracy (66%) and P5 on overall satisfaction (63%). Additionally, there was a positive perception compared to human support, with 78% of participants stating the experience was "Much Better" (41%) or "Better" (37%). Finally, the intention to reuse the tool reached 100%, meaning that all participants found the chatbot useful.

From a practical perspective, these results imply that implementing similar chatbots could significantly reduce administrative bottlenecks and improve the user experience in state health systems by facilitating immediate and accurate access to insurance information. Scientifically, the study contributes evidence that hybrid AI models can successfully merge deterministic and generative reasoning to improve the interpretability and adaptability of automated agents in specialized domains.

Furthermore, the 100% intention to reuse the tool and the 78% positive comparison to human support highlight users' trust in this type of system, reflecting its potential to complement rather than replace human staff. Future research could explore expanding the dataset used for RNN training, integrating reinforcement learning for adaptive responses, and testing the model in other public service areas, thus consolidating AI's role in citizen-oriented digital transformation.

ACKNOWLEDGMENT

Thanks to the Dirección de Investigación of the Universidad Peruana de Ciencias Aplicadas for their support to carry out this research work through the UPC-EXPOST-2025-2 incentive.

ETHICAL APPROVAL

The study involved only anonymous usability testing conducted by voluntary users. No clinical data or personal information were collected, and the study posed no risk to participants. Although formal ethical approval was not required, the research was conducted in accordance with the general ethical principles outlined in the Declaration of Helsinki.

INFORMED CONSENT

When accessing the application, all participants were clearly informed about the purpose of the study and that no personal data was required. Participants voluntarily agreed to take part in the usability testing, and upon entering the system, they enabled the option confirming their understanding of the application's purpose and provided their consent through the platform itself to proceed with its use.

DISCLOSURE STATEMENT

The authors report there are no competing interests to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the Open Science Framework (OSF) at: https://doi.org/10.17605/OSF.IO/F49Q7.

REFERENCES

[1] Superintendencia Nacional de Salud (2022). Registro Nominal de Asegurados. REGINA Boletín Informativo. https://ipe.org.pe/wp-content/uploads/2022/02/SUSALUD-Boletin-REGINA-31-de-enero-2021.pdf

[2] Centro Nacional de Planeamiento Estratégico. (2023). Mayor población afiliada a un sistema de salud. Observatorio Ceplan. https://observatorio.ceplan.gob.pe/ficha/t24

[3] Ramos, A. (2024). Ingresos de EsSalud se duplicaron, pero consultas ambulatorias de afiliados solo crecieron 5% en 15 años - Infobae. Infobae. https://www.infobae.com/peru/2024/01/14/ingresos-de-essalud-se-duplicaron-pero-consultas-ambulatorias-de-afiliados-solo-crecieron-5-en-15-anos/

[4] Tovar, A. (2023). EsSalud: radiografía de la crisis que afecta a 12 millones de asegurados. Salud con Lupa. https://saludconlupa.com/noticias/essalud-radiografia-de-la-crisis-que-afecta-a-12-millones-de-asegurados/

[5] Nuruzzaman, M., & Hussain, O. K. (2020). IntelliBot: A Dialogue-based chatbot for the insurance industry. Knowledge-Based Systems, 196, 105810. https://doi.org/10.1016/j.knosys.2020.105810

[6] Li, Z., Yang, X., Zhou, L., Jia, H., & Li, W. (2023). Text Matching in Insurance Question-Answering Community Based on an Integrated BiLSTM-TextCNN Model Fusing Multi-Feature. Entropy, 25(4), 639. https://doi.org/10.3390/e25040639

[7] Choudhary, P., & Chauhan, S. (2023). An intelligent chatbot design and implementation model using long short-term memory with recurrent neural networks and attention mechanism. Decision Analytics Journal, 9, 100359. https://doi.org/10.1016/j.dajour.2023.100359

[8] Siddiqi, D. A., Miraj, F., Raza, H., Hussain, O. A., Munir, M., Dharma, V. K., Shah, M. T., Habib, A., & Chandir, S. (2024). Development and feasibility testing of an artificially intelligent chatbot to answer immunization-related queries of caregivers in Pakistan: A mixed-methods study. International Journal Of Medical Informatics, 181, 105288. https://doi.org/10.1016/j.ijmedinf.2023.105288

[9] Nguyen-Mau, T., Le, A., Pham, D., & Huynh, V. (2024). An information fusion based approach to context-based fine-tuning of GPT models. Information Fusion, 104, 102202. https://doi.org/10.1016/j.inffus.2023.102202

[10] Gidey, K., Beyene, H., Haileslassie, F. & Berihu, H. (2025). Apprentice bot model design and implementation for psychological clients' therapy. Discover Applied Sciences, 7(3), 170. https://doi.org/10.1007/s42452-025-06515-2

[11] Martín, A. (2018). El contrato de seguro de asistencia sanitaria y la protección del asegurado. Universidad de Almería. http://hdl.handle.net/10835/7228

[12] Gutiérrez, C., Romaní, F. R., Wong, P., & Del Carmen Sara, J. (2018). Brecha entre cobertura poblacional y prestacional en salud: un reto para la reforma de salud en el Perú. Anales de la Facultad de Medicina, 79(1), 65. https://doi.org/10.15381/anales.v79i1.14595

[13] White, L., Fishman, P., Basu, A., Crane, P. K., Larson, E. B., & Coe, N. B. (2019). Medicare expenditures attributable to dementia. Health Services Research, 54(4), 773-781. https://doi.org/10.1111/1475-6773.13134

[14] Mueller, J., Makridis, C. A., Tiffany, T., Borkowski, A. A., Zachary, J., & Alterovitz, G. (2024). From theory to practice: Harmonizing taxonomies of trustworthy AI. Health Policy OPEN, 100128. https://doi.org/10.1016/j.hpopen.2024.100128

[15] Lim, H., Park, Y., Hong, J. H., Yoo, K., & Seo, K. (2024). Use of machine learning techniques for identifying ischemic stroke instead of the rule-based methods: a nationwide population-based study. European Journal Of Medical Research, 29(1). https://doi.org/10.1186/s40001-023-01594-6

[16] Weinberg, S. T., Monsen, C., Lehmann, C. U., Leu, M. G., Webber, E. C., Alexander, G. M., Beyer, E. L., Chung, S. L., Dufendach, K. R., Hamling, A. M., Harper, M. B. & Kirkendall, E. S. (2021). Integrating Web Services/Applications to Improve Pediatric Functionalities in Electronic Health Records. Pediatrics, 148(1). https://doi.org/10.1542/peds.2021-052047

[17] Palta, R., Angwin, J. & Nelson, A. (2024). How We Tested Leading AI Models Performance on Election Queries. The AI Democracy Projects. https://www.proofnews.org/how-we-tested-leading-ai-models-performance-on-election-queries/

[18] Mohammad, M. (2023). An evaluation of the critical success factors impacting artificial intelligence implementation. International Journal of Information Management, 69, 102545. https://doi.org/10.1016/j.ijinfomgt.2022.102545

[19] Erickson, J. (2024). MySQL: qué es y cómo se usa. Oracle. https://www.oracle.com/ar/mysql/what-is-mysql/

[20] Waqas, M. & Wannasingha, U. (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. MethodsX, 13, 102946. https://doi.org/10.1016/j.mex.2024.102946

[21] Islam, S., Elmekki, H., Elsebai, A., Bentahar, J., Drawel, N., Rjoub, G. & Pedrycz, W. (2024). A comprehensive survey on applications of transformers for deep learning tasks. Expert Systems with Applications, 241, 122666. https://doi.org/10.1016/j.eswa.2023.122666

[22] Utomo, A., Wijaya, G., & Setiawan, Y. (2023). Implementation of Crowdfunding Web Application Using AWS Amplify Architecture with End-to-End Testing Using Playwright. Indonesian Journal Of Multidisciplinary Science, 2(11), 3888-3904. https://doi.org/10.55324/ijoms.v2i11.620

[23] Elordi, U., Unzueta, L., Goenetxea, J., Sanchez-Carballido, S., Arganda-Carreras, I., & Otaegui, O. (2020). Benchmarking Deep Neural Network Inference Performance on Serverless Environments With MLPerf. IEEE Software, 38(1), 81-87. https://doi.org/10.1109/ms.2020.3030199

[24] Morgado, J. (2024). Deploy your first web app on AWS with AWS Amplify, Lambda, DynamoDB, and API Gateway. https://dev.to/juliafmorgado/deploy-your-first-web-app-on-aws-with-aws-amplify-lambda-dynamodb-and-api-gateway-2ah7

[25] Mandic, D., Miscevic, G. & Bujišić, L. (2024). Evaluating the Quality of Responses Generated by ChatGPT. Metodicka praksa. 27(1), 5-19. https://doi.org/10.5937/metpra27-51446

APPENDIX A: PSEUDO CODE OF THE PROPOSED RNN–AI MODEL

```
System Initialization:
a. Load the pre-trained Recurrent Neural Network (RNN) model from
"modelo.keras".
b. Load the tokenizer ("tokenizer.pkl") and label encoder
("label_encoder.pkl"), required for text processing and response
interpretation.
c. Define global system parameters:
i. Maximum sequence length: 20
ii. Confidence threshold: 0.7

Query Reception:
a. The system receives the user's input (Q).
b. Validate that the question is not empty and does not contain
unrecognized or invalid characters.

Text Preprocessing and Normalization:
a. Convert the entire query to lowercase.
b. Tokenize the text, converting each word into a numeric index according
to the trained vocabulary.
c. Apply padding to the resulting sequence to reach the predefined maximum
length.

RNN Inference:
a. Feed the processed sequence into the RNN model.
b. The model produces a probability distribution (P) over all possible
responses in its trained set.

Prediction Confidence Evaluation:
a. Identify the maximum probability (prob_max = max(P)) and the associated
index (response_index = argmax(P)).
b. If the obtained probability is greater than or equal to the defined
threshold, decode the response using the label encoder.

Base Response Generation:
a. If (prob_max ≥ confidence_threshold), obtain the preliminary response
(R) from the identified index.
b. If no relevant match is found, set a generic response:
R ← "I'm sorry, I don't have that information."

Response Postprocessing:
a. Clean and format (R), removing duplicates, errors, or unnecessary
characters.

Integration with the Generative Model (OpenAI):
a. Send the base response (R) as input to OpenAI's language model.
b. OpenAI analyzes, reformulates, and enriches the content, producing a
more natural, clear, and contextualized final response: (R_final).

Error and Exception Handling:
a. If an error occurs while loading model files or performing inference,
log the event in the system logs.
b. In such cases, return the message: "The request could not be processed."

System Output:
Present the final response (R_final) to the user as the result of the
chatbot interaction.
```